

A gene trap approach in mouse embryonic stem cells: the *lacZ* reporter is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice

William C. Skarnes,^{1,2} B. Anna Auerbach,³ and Alexandra L. Joyner^{1,3,4}

³Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada, M5G 1X5; ¹Department of Molecular and Medical Genetics, University of Toronto, Toronto, Canada

XP 000569059

We have confirmed that the gene trap vector pGT4.5 creates spliced fusion transcripts with endogenous genes and prevents the synthesis of normal transcripts at the site of integration. cDNA was prepared to the *lacZ* fusion transcript in three ES cell lines to recover endogenous exon sequences upstream of *lacZ*. Each of the clones detected a unique-sized endogenous transcript, as well as the fusion transcript in the ES cell line from which the clone was derived. Sequence analysis of these clones and larger clones isolated from a random-primed cDNA library showed that the splice acceptor was used properly. For two insertions, the expression patterns of the *lacZ* reporter and the associated endogenous gene were compared in situ at three embryonic stages and were found to be similar. Three gene trap insertions were transmitted into the germ line, and abnormalities were observed with two of the three insertions in the homozygous state. RNA obtained from mice homozygous for the two mutant gene trap insertions was analyzed for normal endogenous transcripts and negligible amounts were detected, indicating that little splicing around the gene trap insertion occurred. This work demonstrates the capacity of the gene trap vector to generate *lacZ* fusion transcripts, to accurately report endogenous gene expression, and to mutate the endogenous gene at the site of integration.

[Key Words: Gene trap vector; mouse mutants; embryonic stem cells; *lacZ* reporter; endogenous gene expression]

Received February 3, 1992; revised version accepted March 26, 1992.

A strategy to monitor transcriptionally active regions of the genome was first described in bacteria; it involves the introduction of reporter constructs into the genome that require the acquisition of *cis*-acting DNA sequences to activate reporter gene expression (Casadaban and Cohen 1979). In this way, genes are identified based on expression information and subsequently cloned from DNA sequences flanking the site of insertion. This approach has been applied more recently in higher organisms using modified vectors suitable for eukaryotic transcription units (for review, see Bellen et al. 1990; Skarnes 1990).

One type of vector, termed enhancer traps, was designed to capitalize on the observation that cellular enhancers possess the ability to activate (or repress) transcription over a distance of several kilobase pairs (kbp) independent of their orientation with respect to a gene and are capable of acting upon heterologous promoters

(Fried et al. 1983; Weber et al. 1984; Hamada 1986a, 1986b). This approach provided an efficient method to identify enhancers that activated or repressed transcription of a reporter containing a minimum promoter following differentiation of cultured cell lines (Bhat et al. 1988; Okamoto et al. 1990).

In *Drosophila*, the enhancer trap strategy has been used successfully in large-scale screens for genes expressed at particular developmental stages or in particular lineages (O'Kane and Gehring 1987; Fasano and Kerridge 1988; Fasano et al. 1988; Bellen et al. 1989; Bier et al. 1989; Wilson et al. 1989; for review, see Bellen et al. 1990). Using P-element-based vectors, thousands of insertions have been generated to monitor chromosomal loci active during embryogenesis. The *lacZ* reporter provides a sensitive and easily assayable gene product to detect expression in whole embryos. Approximately 65% of the lines tested were found to express the *lacZ* reporter in a restricted pattern during embryogenesis. Furthermore, ~15% of the P-element insertions caused recessive mutations that resulted in visible phenotypes.

²Present address: Centre for Genome Research, Edinburgh, UK.

*Corresponding author.

In similar studies, one in five transgenic mice carrying enhancer trap vectors were found to exhibit unique temporal and spatial patterns of *lacZ* expression (Allen et al. 1988; Kothary et al. 1988). In one case, the integration event resulted in a mutant phenotype (Kothary et al. 1988). To effectively use this strategy on a large scale in mice, we introduced *lacZ* reporter constructs into mouse embryonic stem (ES) cells (Gossler et al. 1989). Large numbers of integration events can be rapidly generated in ES cells, and the pattern of *lacZ* expression can be easily assayed in ES cell-derived chimeric embryos.

The ability to prescreen ES cells for desired insertion events also made possible the design of a second type of vector that we have termed the gene trap (Gossler et al. 1989). Gene trap vectors were designed to generate spliced fusion transcripts between the reporter gene and the endogenous gene present at the site of integration (Brenner et al. 1989; Gossler et al. 1989; Kerr et al. 1989; Friedrich and Soriano 1991). The essential feature of the gene trap design is the placement of a splice acceptor in front of a promoterless *lacZ* gene. Integrations of the vector into the intron of a gene in the correct orientation were predicted to create *lacZ* fusion transcripts, and if the reading frames of the endogenous genes and *lacZ* are the same, an active β -galactosidase fusion protein should be produced. Vectors similar to the original bacterial vectors, termed promoter traps, have also been used (von Melchner et al. 1990; Hope 1991; Reddy et al. 1991). Because they contain only the coding sequences of the reporter gene, they are expected to require insertions into exons to activate reporter gene expression.

If *lacZ* expression faithfully mimics that of the endogenous gene, monitoring the *lacZ* fusion gene activity in embryos should enable one to readily visualize the pattern of endogenous gene expression during development. The accuracy of *lacZ* expression associated with enhancer/gene/promoter trap insertions has not been reported in the mouse to date. In *Drosophila*, the majority of enhancer trap insertions tested express β -galactosidase in a pattern similar to the adjacent endogenous gene (Fasano et al. 1988; Bellen et al. 1989; Bier et al. 1989; Wilson et al. 1989). The requirement for gene/promoter trap insertions to occur within transcription units may well favor an accurate portrayal of endogenous gene expression.

The introduction of exogenous DNA into the mouse germ line has the potential to generate mutations. Approximately 5% of retrovirus and 10–20% of transgene insertions have been found to cause recessive phenotypes in the mouse (for review, see Jaenisch 1988). The mutated locus can be cloned from genomic DNA flanking the site of transgene insertion. However, the identification of transcription units from flanking sequences requires considerable effort. Consequently, the molecular characterization of genes associated with only one transgene-induced mutation (Maas et al. 1990; Woychik et al. 1990) and three retrovirus-induced mutations (Schnieke et al. 1983; Gridley et al. 1990; Weiher et al. 1990) has been published to date.

Gene and promoter trap vectors may represent more

powerful insertional mutagens in the mouse than other transgenes (Gossler et al. 1989; Friedrich and Soriano 1991). This is predicted from the fact that these vectors can create *lacZ* fusion products with endogenous genes and, as a result, may interfere with the normal coding capacity of the endogenous gene and thereby create a mutation. Moreover, cloning a portion of the endogenous gene directly from the *lacZ* fusion transcript eliminates the time-consuming task of searching for exons in flanking genomic DNA.

Before using a gene trap approach in a large-scale screen to identify novel developmentally regulated genes, it was necessary to validate that the gene trap vector was functioning as predicted. In this study we characterized four ES cell lines that carried independent integrations of a gene trap vector, pGT4.5, and expressed β -galactosidase in ES cells and during embryogenesis. From three cell lines, a portion of the endogenous gene was cloned from the *lacZ* fusion transcripts; and in each case, the splice acceptor in the gene trap vector was used properly. The three endogenous genes identified were novel, one of which encodes a zinc finger-containing protein.

The fidelity of the *lacZ* reporter gene to reflect the expression pattern of the endogenous gene was investigated by comparing the endogenous transcript distribution with the sites of β -galactosidase expression. To directly test the potential of the gene trap insertions to interrupt normal splicing of the endogenous gene, we examined RNA from mice homozygous for two insertions. Splicing around the gene trap insertion was estimated to generate <0.1% of normal endogenous transcript levels in homozygous tissues. Moreover, phenotypic abnormalities were observed for two of three insertions transmitted into the germ line. Our studies with the pGT4.5 vector demonstrate the feasibility of a gene trap approach to identify, mutate, and clone genes important for normal mouse development and physiology.

Results

In this study we concentrated on four pGT4.5 gene trap integrations: Two (Gt2 and Gt4-2) were presumed to be constitutively expressed, and two (Gt4-1 and Gt10) displayed interesting patterns of expression between 8.5 and 12.5 days of development (Gossler et al. 1989). The endogenous genes associated with the Gt4-1, Gt4-2, and Gt10 insertions were cloned (see below), allowing a comparison to be made of the distribution of the endogenous transcripts and *lacZ* activity. The Gt2, Gt4-1, and Gt4-2 ES cell lines were transmitted into the germ line, and the mutant phenotypes associated with these insertions were assessed. *lacZ* expression for the Gt2, Gt4-1, and Gt4-2 inserts was analyzed in these transgenic embryos. For the analysis of the Gt10 *lacZ* expression, ES cell chimeric embryos were generated.

The gene trap vector creates lacZ fusion transcripts with endogenous genes

In our earlier studies, differences in the subcellular lo-

calization of the β -galactosidase activity suggested that the splice acceptor of the gene trap vector pGT4.5 was used to produce unique *lacZ* fusion products (Gossler et al. 1989). To examine whether each gene trap insertion expressed a unique-sized *lacZ* fusion transcript, RNA was purified from seven gene trap cell lines and analyzed by Northern blot hybridization with a *lacZ* probe (Fig. 2, below, and data not shown). The fusion transcripts varied from 3.4 to 10 kb in size, although most (5 of 7) were <4 kb. Because the *lacZ*-containing exon should contribute 3.3 kb, the majority of the fusion transcripts included <0.7 kb of the endogenous gene.

The rapid amplification of cDNA ends (RACE) protocol (Frohman et al. 1988) was used to clone cDNA sequences spanning the *lacZ* splice junction from three cell lines. The amplified products contained 150 bp of the *En-2* sequences downstream of the splice junction. cDNAs generated from RNA from the cell lines Gt4-1 and Gt10 were ~0.5 kb in length, the predicted length of cDNA products that extend to the 5' end of the fusion transcripts. In contrast, a heterogeneous smear of cDNA was amplified from RNA from cell line Gt4-2, in keeping with the estimated 7 kb of endogenous sequence present in the fusion transcript. From each cell line, amplified cDNA ~0.5 kb in size was size selected and cloned.

The gene trap splice acceptor is used properly to create novel β -galactosidase fusion products

At least three independent cDNA clones of different sizes from each cell line were sequenced. Figure 1 shows the nucleotide sequence of the *En-2* splice acceptor used in the gene trap vector and the sequence of the longest amplified cDNA obtained for each cell line. In each cell line, novel sequences were found upstream of the splice junction and none of the clones contained *En-2* intron sequences. As expected, because *lacZ* did not contain a translation initiation codon, an open reading frame (ORF) extended upstream of the splice junction that was in-frame with *lacZ*.

The sequence of the Gt4-1 and Gt10 clones, which contained most of the 5' ends of these genes, had equivalent numbers of CpG and GpC dinucleotides at their 5' ends, suggesting that these two genes are associated with CpG islands (Bird 1986). The Gt4-1 sequence (Fig. 1B) contained a single ORF with an ATG Kozak consensus sequence (Kozak 1987). The 5' end of the Gt10 gene showed evidence of alternative splicing and/or alternate promoter use because two of the six clones contained unique sequences that diverged 312 bp upstream of the splice site and 66 bp upstream of a potential translation initiation signal (Fig. 1D). An amount of 0.4 kb of the Gt4-2 fusion transcript was recovered, and it contained a single ORF compatible with *lacZ* (Fig. 1C).

To further characterize the three endogenous genes, we attempted to obtain larger endogenous cDNA clones that spanned the splice junction by screening a cDNA library prepared from 12.5-day embryonic RNA (see Materials and methods). Three overlapping Gt4-2 cDNA clones that span a total of 5.5 kb and a single 2-kb Gt10

cDNA clone were recovered. The Gt10 cDNA began in the exon sequences common to all forms of the Gt10-amplified cDNA clones and extended 1.7 kb downstream of the *lacZ* splice site. No cDNA clones for the Gt4-1 gene were recovered.

The nucleotide sequences downstream of the splice junction to *lacZ* were determined, and they showed that the reading frames identified in the original amplified cDNA clones continued (Fig. 1C,D). It therefore appears that the Gt4-2 and Gt10 gene trap insertions have interrupted the normal coding sequences of the associated endogenous gene. Taken together, these results demonstrate that the splice acceptor site in the pGT4.5 vector is used correctly.

Amplified cDNA clones detect both the endogenous and fusion transcripts in ES cells

The amplified cDNA clones were used to determine the sizes of the normal uninterrupted endogenous transcripts (Fig. 2; Table 1). Unique probes that hybridized to single bands on genomic Southern blots (data not shown) were used for Northern blot analysis of total ES cell RNA obtained from each cell line (Fig. 2).

The amplified Gt4-2 probe detected a 10-kb endogenous transcript that was similar in size to the Gt4-2 fusion transcript (Fig. 2). A probe containing sequences common to all of the Gt10 cDNA clones hybridized to a major endogenous 2.9-kb transcript and a minor 2.3-kb transcript in all ES cellular RNA and a 3.5-kb fusion transcript in Gt10 RNA. The Gt4-1 probe detected an 8-kb transcript in the Gt4-2 and Gt10 ES cellular RNA but apparently only the 3.4-kb fusion transcript in the Gt4-1 line.

To investigate whether endogenous exons downstream of the gene trap insertion were incorporated into aberrant transcripts that were either spliced abnormally or transcribed from the gene trap vector, the Northern blot was reprobed with Gt10 and Gt4-2 exon probes 3' of the *lacZ* splice site. Only the normal-sized endogenous transcripts were detected (Fig. 2), and, as expected, these sequences were not included in the *lacZ* fusion transcripts.

*Comparison of *lacZ* fusion and endogenous gene expression during development*

A description of the *lacZ* expression pattern during development was obtained from serially sectioned embryos at selected stages and in newborn mice (see Materials and methods). Table 1 summarizes the major sites of β -galactosidase activity during embryogenesis for each gene trap insertion. To examine the distribution of endogenous transcripts, *in situ* hybridizations were performed using RNA probes derived from endogenous Gt4-2 and Gt10 cDNA clones. The short 0.2-kb Gt4-1 antisense probe did not provide conclusive results, probably due to a lack of sensitivity because the *lacZ* fusion product is expressed at very low levels.

Gt10 gene expression Staining of whole-mount 9.5-day and sections of 10.5- and 12.5-day chimeric embryos gen-

B 679-1 3'

C 54-2 5'

5

0 5710 5

Gt4-2 expression From our preliminary analysis of Gt4-2 expression in whole-mount chimeric embryos, we reported that the Gt4-2 fusion gene product was ex-

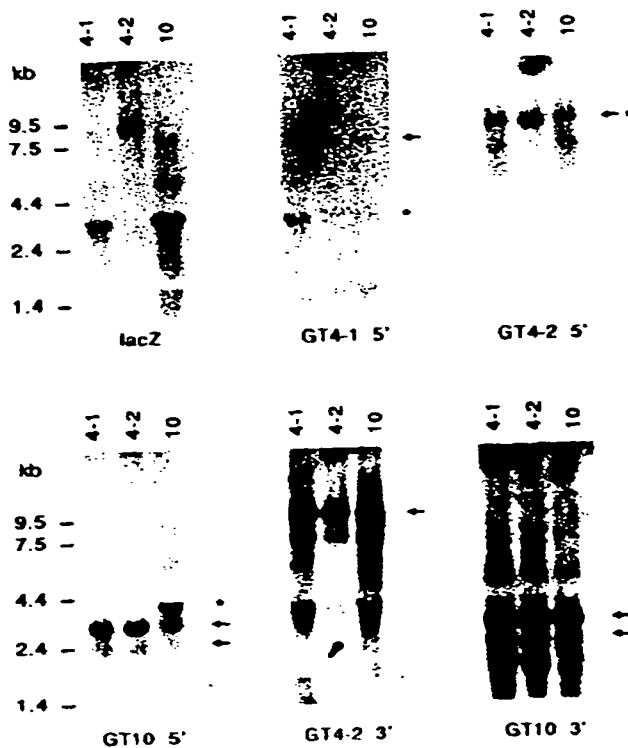


Figure 2. *lacZ* fusion and endogenous transcripts associated with three gene trap cell lines. Northern blot analysis of 5 μ g of total ES cellular RNA from the Gt4-1, Gt4-2, and Gt10 cell lines hybridized with the probes indicated. The same blot was used for each probe. Asterisks (*) show the position of the *lacZ* fusion transcript identified with the *lacZ* probe. The normal endogenous transcripts are indicated by arrows. Autoradiographs with the Gt4-2 and Gt10 3' probes were overexposed to allow for detection of rare abnormal transcripts.

pressed constitutively between 8.5 and 12.5 days of gestation [Gossler et al. 1989]. A more detailed analysis of sections of chimeric and transgenic embryos at these stages showed that *lacZ* expression was absent in a subset of tissues, notably, the yolk sac endoderm, the heart, and the liver. Highest levels of expression were seen in the developing peripheral and central nervous systems [PNS and CNS, respectively] (Fig. 4a,c,e,g). At 15.5 days, *lacZ* expression was restricted to almost all cells of the PNS and CNS. The submandibular gland was one of the few non-neuronal tissues that expressed the reporter gene.

The distribution and relative levels of the endogenous Gt4-2 transcript matched the sites of *lacZ* expression at all stages examined (Fig. 4b,d,f,h). In 9.5- and 12.5-day embryos, the Gt4-2 transcript was distributed widely throughout the embryo, with the highest levels found in the neural tube and little or no signal in the heart and liver (Fig. 4b,d,f). At 15.5 days, expression was strong in all regions of the brain, spinal cord, and dorsal root ganglia, with lower levels in the submandibular gland (Fig. 4h).

Gene trap insertions can cause recessive abnormalities

To test whether the GT4.5 insertions caused phenotypic abnormalities, we produced germ-line chimeric mice from three ES cell lines: Gt2, Gt4-1, and Gt4-2 (data not shown). The Gt10 cell line did not give rise to germ-line chimeras. Agouti offspring from matings of germ-line chimeras were genotyped by Southern blot analysis with a *lacZ* probe, and, as expected, ~50% of the mice contained the gene trap insertion (data not shown).

Offspring heterozygous for the gene trap insertions were backcrossed one generation to allow random segregation of any potential secondary mutations acquired during ES cell culture. The expected 1:1 ratio of wild-type and heterozygous animals were born and reached adulthood with each insertion (Table 2). Heterozygous animals obtained from the backcross were then intercrossed to determine the phenotype of homozygous animals. The breeding results are summarized in Table 2.

Gt4-2 homozygotes die or are growth retarded A restriction fragment length polymorphism (RFLP) caused by the Gt4-2 gene trap insertion and detected with a Gt4-2 5' exon probe (Fig. 5) was used to determine the genotype of Gt4-2 offspring. Roughly, the expected 1:2:1 frequency of wild-type, heterozygous, and homozygous pups was present in a total of 70 pups analyzed at weaning on a mixed genetic background (Table 2). However, homozygous pups were visibly smaller than their wild-type littermates of the same sex at weaning, and differences in growth rates first became apparent at 1–2 weeks of age. Animals were weighed at 4–5 weeks and at 8 weeks: Homozygous animals were 50–75% the weight of their wild-type littermates at 4–5 weeks and near normal (90%) body weight at 8 weeks (data not shown). On the inbred 129/Sv genetic background (Table 2), fewer than the expected number of Gt4-2 homozygous animals were obtained (1 of 31 live offspring) and 3 of the 24 heterozygotes were half the normal size. The phenotypic consequences of the Gt4-2 defect therefore seems to be more severe in the 129/Sv genetic background, causing lethality in the mutant homozygous state and a semidominant effect on growth in the heterozygous state.

Gt4-1 homozygotes die at birth To give an indication of the expression pattern of the endogenous Gt4-1 gene, the pattern of *lacZ* expression was assessed. Transgenic Gt4-1 embryos displayed widespread *lacZ* activity in both extraembryonic and embryonic lineages from 6.5 days to birth. The level of expression varied in different tissues but generally was low in the embryo. In the extraembryonic lineages, particularly in the ectoplacental cone at 8 days (Fig. 6a) and giant cells of 12.5-day placenta (data not shown), expression was high. In the tissues showing the lowest levels, not all of the cells of a given cell type showed staining.

At 8.5 and 9.5 days, staining of whole-mount embryos showed interesting spatial and temporal patterns of β -galactosidase activity. Two stripes of more strongly expressing cells were observed in the developing hindbrain

Table 1. Summary of the characterization of four gene trap cell lines

Gene trap insertion	Subcellular ^a local	Transcript size (kb)		Gene product	Expression ^c pattern	Homozygous phenotype
		<i>lacZ</i> fusion ^b	endogenous			
Gt4-1	cytoplasm (dot)	3.6	8	?	widespread (placenta)	perinatal lethal
Gt4-2	nuclear	10	10	zinc finger protein	restricted (pan-neural)	growth retarded or lethal
Gt10	cytoplasm (vesicles)	3.6	2.9 2.3	proline-rich protein	restricted (otocyst, bv, ^d gut, bladder)	ND
Gt2	nuclear	ND	ND	ND	widespread (neural, kidney, gut)	normal

(ND) Not determined.

^aVisualized by X-gal staining of partially differentiated ES cell monolayers (Gossler et al. 1989).^bGene trap vector sequences contribute ~3.3 kb to the size of the fusion transcript.^cTissues showing highest level of expression are shown in parentheses.^dBlood vessels.

of 8.5- to 9.5-day embryos (Fig. 6b). These stripes appeared to demarcate rhombomeres 3 and 5, similar to *Krox-20* expression at this stage (Wilkinson et al. 1989). At 9.5 days, stronger staining appeared in the otic vesicle and a dorsal/ventral gradient of expression was evident with higher expression on the ventral side of the embryo (data not shown). At 12.5 days and later stages, β -galactosidase activity was found in almost all tissues but at variable levels. The heart showed the highest expression. The mesenchymal cells of the developing lung and gut (Fig. 6c) expressed moderate levels of β -galactosidase activity, whereas the epithelial components of the lung and gut were devoid of expression.

The endogenous Gt4-1 cDNA clone did not detect RFLPs caused by the gene trap insertion with 12 restriction enzymes tested (data not shown). The mice were therefore genotyped with an *En-2* probe by comparing the intensity of the vector band relative to the endogenous *En-2* bands (Fig. 5). In >150 pups analyzed at weaning age from a heterozygous intercross, the expected 1:2 ratio between wild-type and heterozygous pups was found, but there was a complete absence of homozygous animals. Unexpectedly, we identified two viable homozygous animals that survived to adulthood in 22 offspring from a breeding pair that was obtained from a CD-1 (outbred) backcross (Table 2). These findings suggest that there is a strain-dependent penetrance of the lethal phenotype.

We found that the intensity of *lacZ* staining provided a convenient indicator of the genotype of embryos. In litters examined at various stages of development, roughly one-quarter of the embryos did not stain and one-quarter stained more intensely than the remaining embryos. Southern blot analysis confirmed that the most darkly stained embryos were homozygous for the gene trap insertion. Embryos homozygous for the Gt4-1 insertion were found in litters up to the time of birth, and from 17 days to birth these embryos displayed open eyelids. Occasionally, live-born pups with open eyelids were found, which died within a week. A histological examination of serial sections through homozygous and wild-type neonates revealed no obvious defects in the mutants.

The Gt2 insertion has no phenotypic effect Although no sequence data are available on the Gt2 endogenous gene, the fusion product of the Gt2 cell line was found to be chromatin associated (Gossler et al. 1989), suggesting that the Gt2 gene product is a nuclear protein. *lacZ* staining of Gt2 cells could be seen at the metaphase plate of dividing cells, a phenomenon not observed with the nuclear Gt4-2 fusion protein (data not shown). To give an indication of the expression pattern of the endogenous Gt2 gene, the pattern of *lacZ* expression was assessed. Gt2 expression was observed in all tissues at the stages examined (Fig. 6d-f). At 8.5 and 9.5 days, the Gt2 fusion

Figure 3. The β -galactosidase activity and endogenous transcript distribution is similar for the Gt10 gene trap insertion. Shown is comparison of *lacZ* expression in Gt10 chimeric embryos (a,d,g) and the endogenous transcript distribution (b,c,e,f,h,i). *lacZ*-stained 9.5-day whole-mount embryo (a) and a transverse section of a 10.5-day embryo (d), showing high levels of expression in the otic vesicles (o), and caudal (c), and dorsal (d) aortae. No expression was detected in the neural tube (n). A transverse section through a 12.5-day embryo (g) shows β -galactosidase activity in the gut (gl), lung, and dorsal-lateral mesenchyme. Expression was absent in the neural tube and liver (l). In the dark-field micrographs (d,g), *lacZ* staining appears pink. Dark-field photos of in situ hybridization with an endogenous Gt10 probe show transcripts at similar sites in sagittal (b) and transverse (c,e) sections of 9.5-day, sagittal (f) and transverse (h) sections of 12.5-day embryos, and a sagittal section of a 15.5-day embryo (i). The dorsal and caudal aortal, gut, intestinal loops (il), and bladder show the highest levels of hybridization.

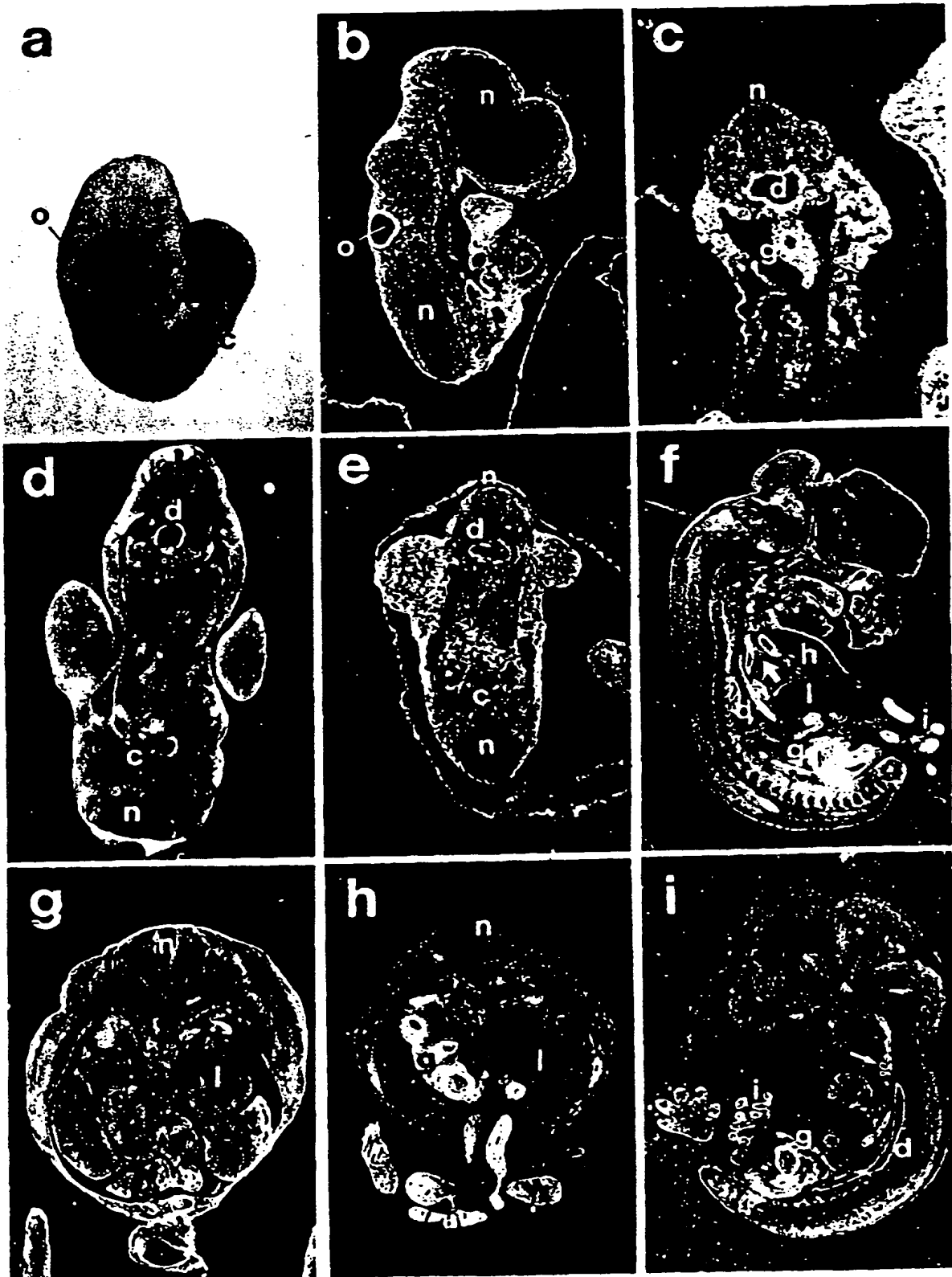


Figure 3. [See facing page for legend.]

product was expressed in all cell types, with the highest level in the developing neural tube (Fig. 6d and data not shown). At 12.5 days and later stages, staining in the forebrain diminished compared with the midbrain and spinal cord (Fig. 6e). High levels of expression were observed in the gut epithelium, in a subset of tubules in the metanephros and mesonephros, and in condensing bone.

Only one transgenic animal was obtained from breeding of the Gt2 germ-line chimera. The mice were genotyped with the *En-2* probe used for the analysis of Gt4-1 mice (Fig. 5). The genomic structure of the Gt2 insertion in transgenic animals was found to differ from that of the original cell line. In the transgenic mice, one of the vector bands present in ES cells in at least two copies was lost and an additional single-copy band was detected (data not shown). The change in restriction pattern is likely the result of a rearrangement that occurred in the chimera mouse. Fertile homozygous pups were present and indistinguishable from their littermates at the expected frequency in a total of 28 offspring (Table 2). Because of variable penetrance of the Gt4-1 and Gt4-2 homozygous phenotype was observed in animals of different genetic backgrounds, the original F₁ Gt2 transgenic animal was backcrossed three times to 129/Sv mice. Heterozygotes were then intercrossed, and among a total of 33 offspring no obvious deviation from the expected Mendelian ratio was observed (Table 2).

The gene trap insertion prevents the synthesis of normally spliced endogenous transcripts

One explanation for the mild phenotypes observed relative to the expression patterns of the three endogenous genes is that the insertions caused leaky mutations owing to splicing around the GT4.5 vector. Ribonuclease protection and reverse transcriptase-polymerase chain reaction (RT-PCR) experiments were used as a sensitive assay to determine whether normal endogenous transcripts were synthesized. RNA was prepared from Gt4-1 and Gt4-2 ES cells and from wild-type, heterozygous, and homozygous Gt4-1 15.5-day embryos or Gt4-2 adult brains. In RNA prepared from homozygous Gt4-1 and Gt4-2 tissues, we did not detect appreciable levels of normally spliced endogenous transcripts (Fig. 7a).

The RT-PCR was used as a more sensitive assay of the amount of normally spliced Gt4-2 transcript in homozygous brain RNA relative to wild-type levels (Fig. 7b). First strand cDNA was prepared from wild-type and homozygous RNA, with a primer specific to Gt4-2 endogenous sequences downstream of the splice site (see Materials and methods). As an internal control, an *En-2* primer specific for the *En-2* and *lacZ* fusion transcripts

was also included in the reverse transcription reaction. Tenfold serial dilutions of the reaction were then subjected to 30 rounds of PCR, with primers designed to amplify normal endogenous Gt4-2 cDNA spanning the splice site, the *lacZ* fusion cDNA, and the *En-2* transcript spanning the splice site. The intensity of the Gt4-2 signal relative to the *En-2* signal in the undiluted cDNA sample from Gt4-2 homozygous RNA was less than the signal obtained with the 10⁻² dilution of wild-type cDNA. The coamplified *En-2* signal showed a similar dilution profile for both samples, indicating that the input RNA and amplification efficiency was similar in both samples. This analysis indicates that <0.1% of the normal transcript was present in brain RNA from homozygous mice. Thus, in the adult brain, where the Gt4-2 gene is expressed at high levels, negligible splicing around the gene trap sequences occurred.

Discussion

Our results confirm the prediction that the pGt4.5 gene trap vector can generate *lacZ* fusion transcripts with endogenous genes through the use of the splice acceptor site contained in the vector. The RACE protocol provided a direct method to clone cDNA sequences present upstream of *lacZ* in the fusion transcripts. The nucleotide sequences of these and additional cDNA clones demonstrated that the splice acceptor in the vector was used to join novel sequences to the 5' end of *lacZ* (Fig. 1). Furthermore, each cDNA probe containing sequences 5' of the *lacZ* splice site detected endogenous transcripts in ES cells, as well as a *lacZ* fusion transcript in the cell line from which the probe was derived (Fig. 2). This result provided definitive evidence that the endogenous gene interrupted by the gene trap insertion had been cloned.

Integrations that activate *lacZ* may occur more often in large transcription units consisting of large introns. The relatively large size of the endogenous transcripts associated with the three gene trap integrations (3, 8, and 10 kb) may indicate such a preference. Southern blot analyses with short endogenous cDNA probes upstream of the insertion site detected more than one fragment when DNA was digested with enzymes that do not cut within the probes (data not shown). This result indicates that the 5' cDNA clones are composed of more than one exon. Thus, the insertions do not appear to show a preference for the first intron, as may be the case with retroviral insertions. The bias toward relatively short fusions may instead reflect the fact that the introns at the beginning of vertebrate genes tend to be much larger than introns farther 3' (Hawkins 1988).

A search of the GenBank data base using both the nu-

Figure 4. *lacZ* expression in Gt4-2 transgenic embryos (a,c,e,g) compared with Gt4-2 endogenous transcript distribution (b,d,f,h). In transverse sections of 9.5-day embryos (a), sagittal (c) and transverse (e) sections of 12.5-day embryos, and a sagittal section of a 15.5-day (g) embryo, β -galactosidase activity was high in the neural tube (n), primitive gut (g), and pharyngeal arches (m). No expression was detected in the heart (h). This pattern matches well with the tissues containing Gt4-2 endogenous transcripts. A comparison of similar sagittal and transverse sections of 9.5-day (b), 12.5-day (d,f) and 15.5-day (h) embryos shows the same pattern in which expression is highest in the CNS and PNS with an absence of expression in the heart and liver.

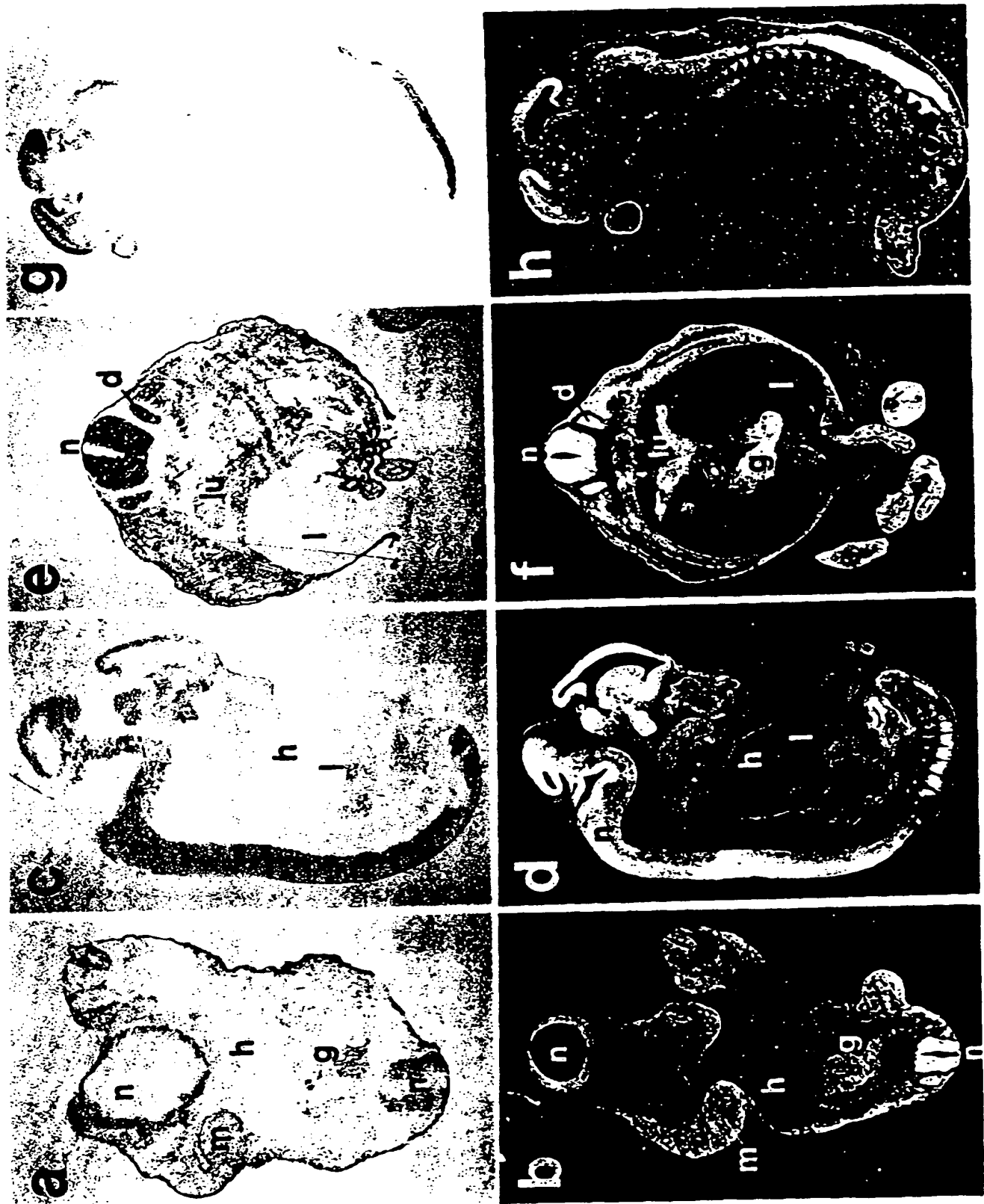


Figure 4. (See facing page for legend.)

Table 2. Summary of transgenic breeding

Breeding pair	Genetic strain ^a	3-4 w.o. pups			Embryos ^b		
		+ / +	+ / -	- / -	+ / +	+ / -	- / -
Gt4-2/+ × + / +	[129 × C57]F ₁ × C57	48	43	NA			
Gt4-2/+ × Gt4-2/+	[129 × C57]F ₁ × C57	26	29	16			
Gt4-2/+ × Gt4-2/+	129	7	24 ^c	2 ^d			
Gt4-1/+ × + / +	[129 × C57]F ₁ × C57	47	62	NA			
Gt4-1/+ × Gt4-1/+	[129 × C57]F ₁ × C57	47	108	0	24	48	21
Gt4-1/+ × Gt4-1/+	[129 × C57]F ₁ × CD1	11	9	2 ^e			
Gt2/+ × + / +	[129 × CD1]F ₁ × CD1	51	61	NA			
Gt2/+ × Gt2/+	[129 × CD1]F ₁ × CD1	7	15	6			
Gt2/+ × Gt2/+	[129 × CD1]F ₁ × 129/B3	9	16	8			

(NA) Not applicable.

^a(C57) C57BL/6; (129) 129/Sv; (CD-1) outbred.^bBased on intensity of β-galactosidase staining and/or DNA analysis.^cTwo heterozygotes that were half the normal weight died.^dOne homozygote that was half the normal weight died.^eOne female, infertile; one male, fertile, homozygous genotype confirmed in backcrosses to CD-1 females.

cleotide and putative protein-coding sequences indicated that the endogenous genes associated with all three gene trap lines were novel. The ORF of the Gt10 gene encodes an amino-terminal proline-rich domain (Fig. 1D) but shows <30% nucleotide sequence identity with other proline-rich proteins in the data base. Three C2H2 zinc finger-coding motifs and an acidic region were present within the Gt4-2 endogenous gene immediately upstream of the *lacZ* splice site (Fig. 1C). The zinc finger motifs share only 30% amino acid identity and 40% nucleotide sequence identity with other members of this gene family. The identity of the Gt4-2 gene as a potential transcription factor readily explains the nuclear localization of the β-galactosidase fusion product (Gossler et al. 1989).

The activation of the *lacZ* reporter in the pGT4.5 gene trap vector requires insertion into a transcription unit and should be dependent on the enhancer, promoter, and translational signals of the endogenous locus. For the Gt4-2 and Gt10 insertions, the distribution and relative abundance of the endogenous transcript matched well the distribution of β-galactosidase activity (Figs. 3 and 4). We expect that some integration events will alter the normal regulatory mechanisms of the endogenous gene. The Gt10 endogenous transcript appeared to be expressed more broadly than *lacZ* in 9.5-day embryos. These minor differences may reflect normal post-transcriptional regulation of the endogenous gene product or, as a consequence of the insertion, a perturbation in the normal transcriptional and/or post-transcriptional control signals. A direct comparison of either endogenous and fusion transcript or protein levels will be required to distinguish between normal regulatory mechanisms or abnormal fusion gene expression. Nevertheless, our results suggest that in most cases the *lacZ* expression pattern will reflect that of the endogenous gene.

The unique patterns of *lacZ* expression observed with each gene trap insertion and the close correlation with

endogenous gene expression for the two lines analyzed implies that the pGT4.5 vector does not contain intrinsic enhancer activity. *lacZ* expression was not found in tissues devoid of endogenous transcripts, nor were obvious patterns of *lacZ* expression found that were common to all insertion events. Therefore, sequences within the vector do not appear to be capable of activating *lacZ* expression outside of the normal domains of endogenous gene expression.

lacZ staining provides a convenient, high-resolution assay to assess the overall pattern of endogenous gene expression. Furthermore, *lacZ* expression should aid in

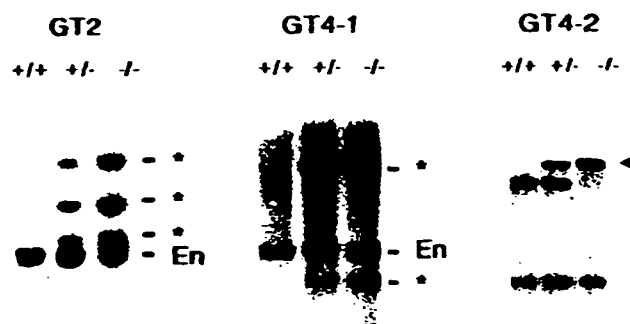


Figure 5. Southern blot analysis to determine the genotype of transgenic mice. *Sst*I (Gt2)- or *Bgl*II (Gt4-1)-digested DNA from tails or embryos was hybridized to the E5 probe (see Materials and methods). The E5 probe detects both transgene (asterisks) and endogenous *En-2* (*En*) fragments. The relative intensities of the transgene fragments compared to the endogenous *En-2* fragments was used to determine the genotype of nontransgenic (+ / +), heterozygous (+ / -), or homozygous (- / -) mutant animals. The 19-A-1 probe (see Materials and methods), containing endogenous Gt4-2 cDNA sequences, detected a new 20-kb *Pst*I fragment (arrow) associated with the gene trap insertion.

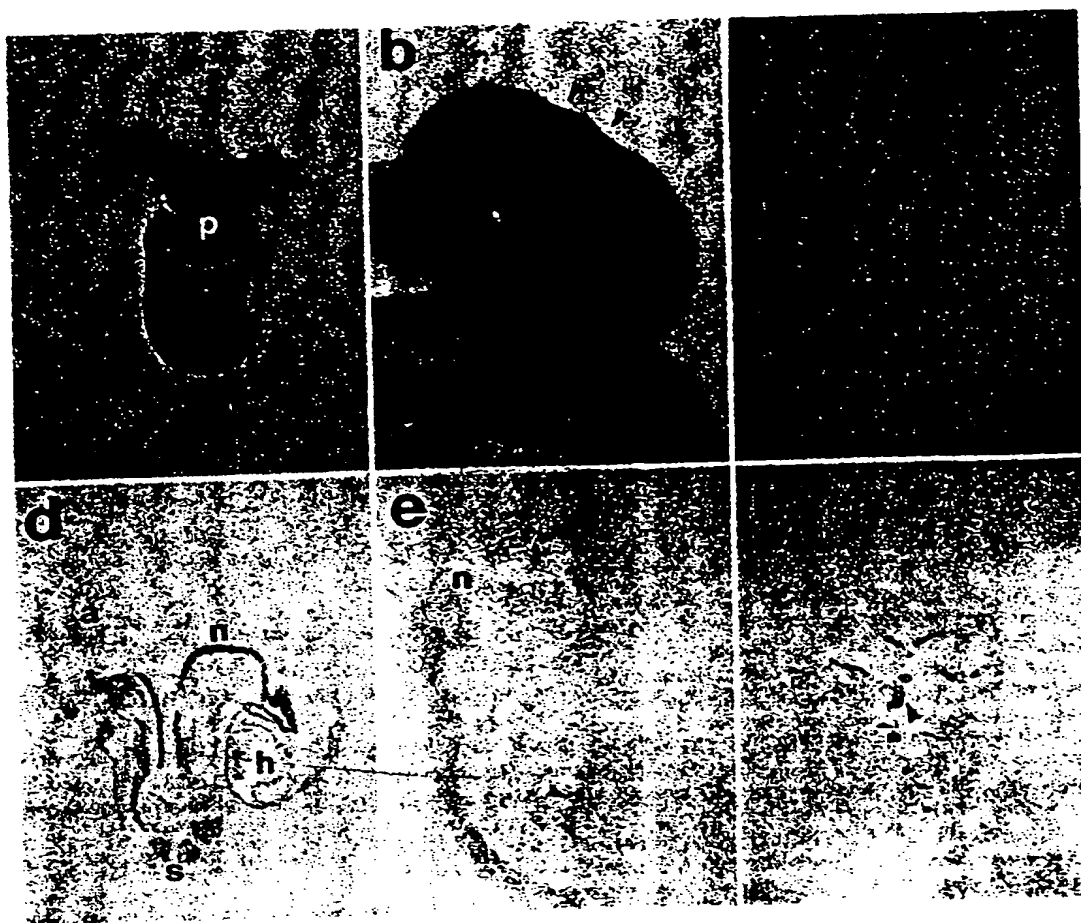


Figure 6. Gt4-1 and Gt2 endogenous transcripts are distributed widely. (a) An 8-day embryo, showing that Gt4-1 *lacZ* expression was highest in the ectoplacental cone. (b) A 9.5-day Gt4-1 embryo, showing high levels of *lacZ* expression in two rhombomeres (indicated by the arrowheads). (c) A section through 12.5-day Gt4-1 intestine, showing one of the few sites where expression of *lacZ* was absent. (d) Sagittal section of an 8.5-day Gt2 embryo, showing high *lacZ* expression in the developing neural tube, (n), moderate levels in mesoderm, that is, somites (s), allantois (a), and heart (h). (e) A mid-sagittal section of a 12.5-day embryo, showing the highest expression of Gt2 *lacZ* in the neural tube (n), developing gut (arrowhead), and mesonephros. (f) A higher magnification of a 15.5-day Gt2 kidney, showing high expression in the proximal tubules (arrowhead).

the phenotypic analysis of gene trap mutations. First, β -galactosidase activity marks the tissues that normally express the mutated gene. Defects in cells that express *lacZ* would indicate a cell-autonomous gene function, whereas defects in nonexpressing cells may suggest a cell-nonautonomous role. Second, differences in the patterns of expression in heterozygous versus homozygous embryos may provide important clues regarding the underlying defects caused by the mutation. Finally, *lacZ* staining intensity can offer an easy way to genotype embryos. A comparison of *lacZ*-stained embryos allowed us to quickly pinpoint the stage at which the embryonic lethal mutation associated with the Gt4-1 insertion occurred. This should be of general applicability, particularly for those mutations that fail during preimplantation or early postimplantation development.

The phenotypes associated with the three gene trap

insertions are difficult to equate with the patterns of endogenous gene expression. Homozygous Gt4-1 mice die at birth and display an open eyelid phenotype. The perinatal lethal phenotype of the Gt4-1 mutation is fully penetrant in the C57BL/6 \times 129/Sv hybrid background. However, viable homozygous mice were obtained in a complex genetic background that included both outbred and inbred chromosomes. The location of the Gt4-1 insertion on chromosome 16 [N.A. Jenkins, D.J. Gilbert, and N.G. Copeland, unpubl.] does not correlate with any of the open eye mutants that have been mapped. *lacZ* expression in Gt4-1 mice was widespread, with intriguing stripes of comparatively high expression in rhombomeres at 8.5 days and a dorsal-ventral gradient of expression at 9.5 days. The highest level of expression was found in the placenta. On the surface, the perinatal lethal phenotype seems inconsistent with the pattern of

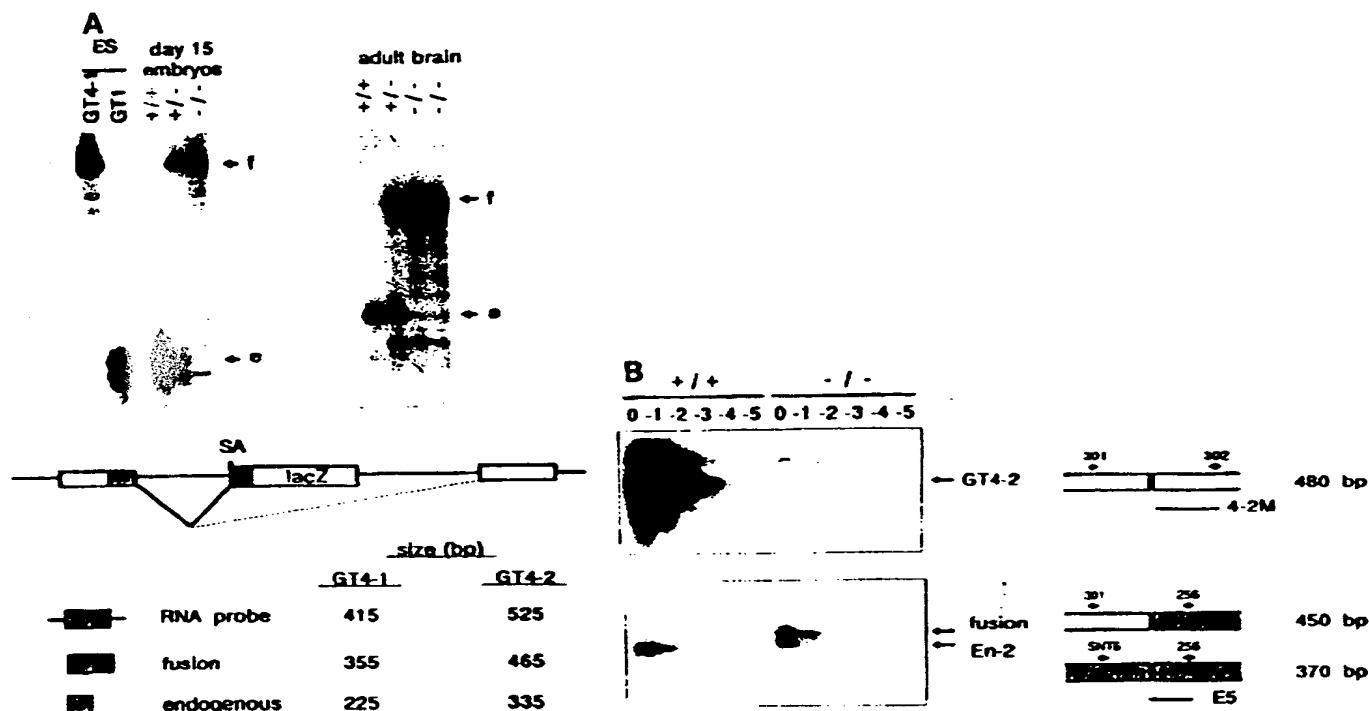


Figure 7. The synthesis of normal endogenous transcripts is negligible in mice homozygous for two gene trap insertions. **[A]** Ribonuclease protection analysis to detect the fusion and endogenous transcript levels in Gt4-1 and Gt4-2 ES cells and 15.5-day Gt4-1 embryos (*left*) and Gt4-2 adult brain RNA (*right*). The protected RNA species corresponding to the fusion (*f*) and the endogenous (*e*) transcripts are indicated. **[B]** PCR amplification of normal Gt4-2 transcripts from wild-type and homozygous Gt4-2 adult brain RNA. Southern blot is shown on the *left* of the PCR products amplified from serial 10-fold dilutions (0–5) of cDNA prepared from (+/+) and (-/-) RNA and detected with the 4-2M probe (which detects the endogenous Gt4-2 transcript) and E5 (which detects the En-2 and lacZ fusion transcripts) probes.

gene expression at these early stages; however, it is possible that subtle abnormalities arose as a consequence of defects manifested early in development.

Homozygous Gt4-2 mice on a C57BL/6 × 129/Sv background showed a mild growth retardation after birth. However, on an inbred 129/Sv background, fewer than the expected number of homozygous mutants survived to weaning and some heterozygotes showed a growth defect. The Gt4-2 endogenous gene is expressed widely during gastrulation and specifically in the CNS and PNS later in development. We speculate that the growth deficiency and lethality may involve a neurological disorder and/or an endocrine defect. In any case, it seems likely that only a subset of cells expressing the gene are responsible for the mutant phenotype. The Gt4-2 gene product contains a zinc finger domain and maps to chromosome 4 (N.A. Jenkins, D.J. Gilbert, and N.G. Copeland, unpubl.), proximal to the major urinary protein (*Mup-1*). The recessive neurological mutants *vacillans* (*vc*) and *whirler* (*wi*) map to this region (summarized in Green 1989). Homozygous *vc* and *wi* animals are smaller than normal and exhibit neurological disorders. The defect in Gt4-2 homozygous mice and the expression of the Gt4-2 gene in the CNS and PNS raise the possibility that the Gt4-2 mutation is an allele of *vc* or *wi*.

Mice that were homozygous for the Gt2 mutation were viable and fertile. The nuclear fusion product was distributed widely throughout development. Sections through homozygous and wild-type neonates showed no gross abnormalities despite high levels of expression in the kidney and gut.

Despite the broad expression patterns of the endogenous genes during the early stages of development, none of the mutations caused early embryonic lethal phenotypes. One explanation to account for this finding is that the gene trap mutation does not completely abolish normal gene activity. One possibility is that the insertions might create leaky mutations owing to partial splicing around the vectors. However, we have shown that negligible amounts of the normal transcripts are present in RNA obtained from homozygous mutant Gt4-1 and Gt4-2 tissues. Furthermore, the insertions do not induce the synthesis of aberrant transcripts other than the fusion transcripts. Therefore, our results indicate that the splice acceptor and poly(A) signals of the pGT4.5 vector can act efficiently to define the final exon of the fusion transcript and prevent splicing events around the insertion. Other gene trap vectors, depending on the individual properties of the splice acceptor and poly(A) signal, may not yield the same results.

The fact that some insertions will generate sizable

β -galactosidase fusion proteins raises the possibility that these fusion products, in some cases, will have partial activity. Alternatively, they may behave as dominant-negative molecules. In this respect, it is interesting that the Gt4-2 mutation may cause a semidominant phenotype, as a growth defect was observed in some heterozygous animals. One interpretation is that the Gt4-2 zinc finger-containing fusion protein has a dominant-negative activity. The more severe phenotype observed in homozygous animals may be a result of the higher levels of fusion protein activity found in these animals. It is also possible that the semidominant phenotype is caused by a dosage effect owing to the loss of gene activity. As with all mutagenesis screens, the severity and nature of individual gene trap mutations may not be immediately obvious and must be examined carefully. The potential exists for pGT4.5 insertions to create dominant-negative, partial loss-of-function and gain-of-function mutations, and not just null mutations. This may be the case for other gene trap vectors as well as for promoter traps.

Mutagenesis screens in a variety of organisms, including the mouse, suggest that the vast majority of genes in the genome are apparently dispensable (Dove 1987). In yeast, where more than half the genome is transcribed, 70% of insertional mutations generated in randomly selected regions of genomic DNA have no detectable phenotype (Goebel and Petes 1986). Similarly, based on a screen for chemically induced recessive phenotypes in the *t* locus of the mouse, it was estimated that 50–100 essential genes reside within this region that represents 1% of the mouse genome (Shedlovsky et al. 1988). This number is an order of magnitude smaller than the number of genes predicted to occupy this region (Ohno 1986). Recently, 24 lines of mice that contain insertions of a retrovirus-based gene trap vector have been generated, and 9 were found to cause recessive embryonic lethal phenotypes (Friedrich and Soriano 1991). We have found one embryonic lethal mutation (Gt4-1) in three lines tested, and a second mutation (Gt4-2) appears to be lethal in inbred mice. Taken together, these results are in line with the prediction that most single gene lesions are not expected to produce obvious phenotypes. Because insertions that likely interrupt an endogenous gene can be preselected in ES cells, it is possible with this approach to recover mutations that are phenotypically silent.

The gene trap approach we describe here combines an effective DNA mutagen with a strategy, similar to the enhancer trap, to identify genes based on patterns of gene expression. This strategy offers a powerful tool for cataloging the expression patterns of genes active during mouse embryogenesis and for creating insertional mutations that are immediately accessible to molecular characterization.

Materials and methods

DNA and RNA probes

For the Northern and Southern blot analyses the endogenous probes that contained sequences 5' of the *lacZ* splice site were

subcloned from the PCR-amplified cDNA products into pGEM-1 (Promega). The clones consisted of the following nucleotides [as shown in Fig. 1]: Gt4-1 5' (bp 72–308), Gt4-2 5' (bp 107–302), Gt10 5' (bp 149–367). Probes that include endogenous sequences spanning the splice site, Gt4-2 3' [0.6 kbp *EcoRI*] 19-A-1 [a 1.1-kb *NcoI*–*KpnI* Gt4-2 fragment], and Gt10 3' [0.5-kbp *XhoI*–*SstI*], were cloned from cDNAs obtained in the screen of a 12.5-day cDNA library. The *En-2* probe used for genotyping transgenics (E5) was a 0.4-kb *BamHI*–*SstI* fragment that includes 0.25 kb of intron and 0.15 kb of exon sequence spanning the *En-2* splice acceptor. The antisense RNA probes used in the RNase protection were cloned into pGEM-7Zf (Promega) and contained the following amplified cDNA sequences (3' → 5') spanning the *lacZ* splice site: Gt4-1a, 130 bp of the *En-2* exon 5' from the *KpnI* site to the splice site (Fig. 1A), 225 bp of Gt4-1 cDNA upstream of the splice site, and 60 bp of the pGEM7Zf polylinker; Gt4-2a, the same 130 bp of *En-2* exon, 335 bp of Gt4-2 cDNA, and 60 bp of the pGEM7Zf polylinker. The 4-2M probe used in the PCR amplification experiment contained 145 bp immediately 3' of the *En-2* splice site.

RNA and DNA blot analyses

Total RNA was prepared by lysis of ES cells or tissues in guanidinium isothiocyanate followed by centrifugation through a cesium chloride cushion (Sambrook et al. 1989). RNA pellets were resuspended in 8 M urea, extracted several times with phenol-chloroform (1 : 1), and stored as an ethanol precipitate. Genomic DNA was purified from proteinase K-digested ES cell pellets and tail biopsies by phenol-chloroform (1 : 1) extraction. Northern and Southern blots using GeneScreen (Dupont) were hybridized with random-primed DNA probes as described (Joyner et al. 1985).

Cloning and sequencing endogenous cDNAs

The RACE strategy (Frohman et al. 1988) with several modifications was used to amplify cDNA sequences 5' to *lacZ*. Five micrograms of total ES cell RNA was annealed with 10 ng of primer 170 (Fig. 1A), and first-strand cDNA was synthesized with reverse transcriptase as described previously (Frohman et al. 1988). Prior to A-tailing, RNA in the reverse transcription reaction was hydrolyzed in alkali [0.2 M NaOH for 1 hr at 65°C]. The single-stranded cDNA was then purified on a NACS column (Bethesda Research Laboratories). A-tailing of the cDNA with terminal deoxytransferase (Boehringer Mannheim) was carried out at 37°C for 7.5 min in the presence of 0.2 mM dATP as described previously (Frohman et al. 1988). The tailed products were extracted once with phenol and once with phenol-chloroform (1 : 1) and precipitated in ethanol. The Klenow enzyme (Pharmacia) was used to synthesize second-strand cDNA in buffer containing 10 ng of primer 210T (5'-GGTTGTGTC-GACTATCGATGGGTTTTTTTTTTTTTTT-3'). The reaction was incubated at room temperature for 30 min and then at 37°C for 30 min. The double-stranded cDNA products were then subjected to 40 rounds of PCR [94°C for 90 sec; 65°C for 2 min; 72°C for 15 min] using 3 units of *Taq* polymerase in 1 × *Taq* buffer (Perkin-Elmer Cetus) containing 1 µg each of primer 210 (5'-GGTTGTGTCGACTATCGATGGG-3') and primer 256 (Fig. 1A). Fresh dNTPs and 1.5 units of *Taq* polymerase were added into the final round of PCR.

The size range of the amplified products was visualized by Southern blot analysis, using the *En-2* ES probe. Size-selected 0.5-kb cDNA was digested with the restriction enzymes *SalI* and *Clal* and cloned into pUC18. Mini-prep plasmid DNA was sequenced with the Sequenase kit (U.S. Biochemical). To obtain

sequence from both strands, exonuclease III deletions were generated (Henikoff 1987) or specific primers were synthesized.

cDNA library screen

A λ gt10 random-primed cDNA library, kindly provided by Mark Hanks (Mount Sinai Hospital, Toronto), was constructed using an Amersham kit from poly(A)⁺-selected 12.5-day embryonic CD-1 mouse RNA. The library contained 2×10^6 inserts and was amplified once. Screens of 2×10^6 phage with the Gt4-1 5' probe, 1×10^6 phage with the Gt4-2 5' probe, and 0.5×10^6 phage with the Gt10 5' probe yielded 0, 3, and 1 clone, respectively, with unique-sized cDNA inserts. The inserts were subcloned into pGEM-7Zf and partially sequenced to confirm their identity.

RNase protection assay

³²P-Labeled antisense runoff transcripts were synthesized with SP6 or T7 polymerase using the Riboprobe kit (Promega). The reaction was treated with DNase I, and the probes were purified on 5% denaturing polyacrylamide sequencing gels. Two to five micrograms of total RNA was hybridized to 10^5 cpm of gel-purified antisense probe and digested with RNase A and T1 as described previously (Melton et al. 1984). The protected RNA species were separated using denaturing polyacrylamide gel electrophoresis and visualized by autoradiography.

PCR amplification of Gt4-2 transcripts

In a volume of 20 μ l, 10 ng of primer 302 (Fig. 1C) and 10 ng of primer 256 were used to prepare Gt4-2- and En-2-specific cDNAs from 2 μ g total RNA. Serial 10-fold dilutions of the cDNA products were subjected to 30 rounds of PCR in 50 μ l of buffer containing 1 μ g each of primer 301 (Fig. 1C), 302, 256, and an En-2 5' primer, SNT6. The 301/302 primers amplify the normal endogenous Gt4-2 transcript to produce a 480-bp cDNA fragment; the 302/256 primer combination amplifies a 450-bp cDNA fragment of the Gt4-2/*lacZ* fusion transcript; and the SNT6/256 primers amplify the endogenous En-2 transcript to produce a 370-bp cDNA fragment containing 240 bp of exon 1 and 130 bp of exon 2.

Production of ES cell chimeras

C57BL/6 blastocysts were injected with 10–15 ES cells (D3) as described previously (Gossler et al. 1989). In one experiment with the Gt2 cell line, CD-1 blastocysts after cell injection were delayed for 5 days in ovariectomized female recipients and then transferred to a new pseudopregnant female. Chimeric males, identified by coat color, were mated to either C57BL/6J (Gt4-1, Gt4-2, and Gt10 lines), or CD-1 (Gt2) or 129/Sv (Gt4-2) females.

lacZ staining of whole-mount and embryo sections

Whole-mount embryos up to 12.5 days were fixed in 0.2% glutaraldehyde (Sigma) and stained with X-gal (Bethesda Research Laboratories) as described previously (Beddington et al. 1989). Beyond 12.5 days, we found that staining of whole-mount embryos was ineffective owing to poor penetration of the stain. Later stage embryos were rinsed in PBS and then frozen slowly to -70°C in O.T.C. embedding medium (Tissue-Tek, Miles, Inc.). The embryos were stored frozen for up to 8 weeks prior to

cryostat sectioning. Frozen sections were fixed in 0.2% glutaraldehyde, 2% formaldehyde, for 5 min, rinsed, and stained.

In situ hybridization

cDNA fragments were cloned into pGEM-7Zf (Promega). The two Gt4-2 probes used were Gt4-2 5' and 19-A-1. The Gt10S (sense) and Gt10X (antisense) probes contained a 1.1-kbp *Xho*I–*Sst*I subfragment of the cDNA clone that spanned the splice site of the Gt10 endogenous cDNA. Each probe was nonrepetitive by Southern blot analysis of genomic DNA and detected a single major transcript by Northern analysis of ES cell RNA.

Uniformly labeled ³⁵S-labeled RNA probes were synthesized with SP6 or T7 polymerase using a Riboprobe kit (Promega). CD-1 embryos were cryostat sectioned at 7 μ m thickness and put onto aminopropyltriethoxysilane-coated slides, fixed in 20% paraformaldehyde, dehydrated, and stored at -70°C . The slides were processed, hybridized under Gelbond coverslips (FMC), and washed as described previously (Hogan et al. 1986). The $2\times$ SSC and $0.1\times$ SSC washes were carried out at 50°C . The slides were then dehydrated and dipped in Kodak NTB-2 emulsion. Exposure times were between 1 and 4 days. At least two separate in situ experiments were carried out with each probe at each stage. Multiple 9.5-day embryos and at least two 12.5- and 15.5-day embryos were analyzed. Both sagittal sections and transverse sections through the head and trunk of 12.5- and 15.5-day embryos were analyzed.

Note added in proof

Sequence data described in this paper have been submitted to the EMBL/GenBank data libraries.

Acknowledgments

We thank Janet Rossant and Martin Breitman for many helpful discussions. We also thank Mark Hanks for the 12.5-day embryonic cDNA library and D. Nallainathan, M. Kownacka, and A. Pearson for excellent technical help. This research was supported by grants from the National Institutes of Health (HD25334) and Bristol-Myers Squibb to A.J. W.C.S. was supported by a [National Cancer Institute of Canada] Steve Fonyo studentship and A.L.J. is a Medical Research Council scholar and an International Research Scholar of the Howard Hughes Medical Institute.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Allen, N.D., D.G. Cran, S.C. Barton, S. Hettle, W. Reik, and M.A. Surani. 1988. Transgenes as probes for active chromosomal domains in mouse development. *Nature* 333: 852–855.
- Beddington, R.S.P., J. Morgerstern, H. Land, and A. Hogan. 1989. An *in situ* transgenic enzyme marker for the mid-gestation mouse embryo and the visualization of inner cell mass clones during early organogenesis. *Development* 106: 37–46.
- Bellen, H.J., C.J. O'Kane, C. Wilson, U. Grossniklaus, R.K. Pearson, and W.J. Gehring. 1989. P-element-mediated enhancer

- detection: A versatile method to study development in *Drosophila*. *Genes & Dev.* 3: 1288-1300.
- Bellen, H., C. Wilson, and W.J. Gehring. 1990. Dissecting the complexity of the nervous system by enhancer detection. *BioEssays* 12: 199-204.
- Bhat, K., M.W. McBurney, and H. Hamada. 1988. Functional cloning of mouse chromosomal loci specifically active in embryonal carcinoma stem cells. *Mol. Cell. Biol.* 8: 3251-3259.
- Bier, E., H. Vaessin, S. Shepherd, K. Lee, K. McCall, S. Barbel, L. Ackerman, R. Carretto, T. Uemura, E. Grell, L.Y. Jan, and Y.N. Jan. 1989. Searching for pattern and mutation in the *Drosophila* genome with a P-lacZ vector. *Genes & Dev.* 3: 1273-1287.
- Bird, A. 1986. CpG-rich islands and the function of DNA methylation. *Nature* 321: 209-213.
- Brenner, D.G., S. Lin-Chao, and S.N. Cohen. 1989. Analysis of mammalian cell genetic regulation in situ by using retrovirus-derived "portable exons" carrying *Escherichia coli lacZ* gene. *Proc. Natl. Acad. Sci.* 86: 5517-5521.
- Casadaban, M.J. and S.N. Cohen. 1979. Lactose genes fused to exogenous promoters in one step using a Mu-lac bacteriophage: In vivo probe for transcriptional control sequences. *Proc. Natl. Acad. Sci.* 76: 4530-4533.
- Dove, W.F. 1987. Molecular genetics of *Mus musculus*: Point mutagenesis and millimorgans. *Genetics* 116: 5-8.
- Fasano, L. and S. Kerridge. 1988. Monitoring positional information during oogenesis in adult *Drosophila*. *Development* 104: 245-253.
- Fasano, L., N. Core, and S. Kerridge. 1988. Expression of a reporter gene resembles that of its neighbour: An insertion in the hairy gene of *Drosophila*. *Wilhelm Roux's Arch. Dev. Biol.* 197: 507-512.
- Fried, M., M. Griffiths, B. Davies, G. Bjursell, G. La Mantia, and L. Lania. 1983. Isolation of cellular DNA sequences that allow expression of adjacent genes. *Proc. Natl. Acad. Sci.* 80: 2117-2121.
- Friedrich, G. and P. Soriano. 1991. Promoter traps in embryonic stem cells: A genetic screen to identify and mutate developmental genes in mice. *Genes & Dev.* 5: 1513-1523.
- Frohman, M.A., M.K. Dush, and G.R. Martin. 1988. Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci.* 85: 8998-9002.
- Goebel, M.G. and T.D. Petes. 1986. Most of the yeast genomic sequences are not essential for cell growth and division. *Cell* 46: 983-992.
- Gossler A., A.L. Joyner, J. Rossant, and W.C. Skarnes. 1989. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* 244: 463-465.
- Green, M.C. 1989. Catalog of mutant genes and polymorphic loci. In *Genetic variants and strains of the laboratory mouse* (ed. M.F. Lyon and A.G. Searle), pp. 12-403. Oxford University Press, Oxford, UK.
- Gridley, T., D.A. Gray, T. Orr-Weaver, P. Soriano, D.E. Barton, U. Francke, and R. Jaenisch. 1990. Molecular analysis of the *Mov 34* mutation: Transcript disrupted by proviral integration in mice is conserved in *Drosophila*. *Development* 109: 235-242.
- Hamada, H. 1986a. Activation of an enhancerless gene by chromosomal integration. *Mol. Cell. Biol.* 6: 4179-4184.
- . 1986b. Random isolation of gene activator elements from human genome. *Mol. Cell. Biol.* 6: 4185-4194.
- Hawkins, J.D. 1988. A survey of intron and exon lengths. *Nucleic Acids Res.* 16: 9893-9909.
- Henikoff, S. 1987. Unidirectional digestion with exonuclease III in DNA sequence analysis. *Methods Enzymol.* 155: 157-165.
- Hogan, B., F. Constantini, and E. Lacy. 1986. *Manipulating the mouse embryo: A laboratory manual*. Cold Spring Laboratory, Cold Spring Harbor, New York.
- Hope, I.A. 1991. "Promoter trapping" in *Caenorhabditis elegans*. *Development* 113: 399-408.
- Jaenisch, R. 1988. Transgenic animals. *Science* 240: 1468-1474.
- Joyner, A.L., T. Kornberg, K.G. Coleman, D.R. Cox, and G.R. Martin. 1985. Expression during embryogenesis of a mouse gene with sequence homology to the *Drosophila engrailed* gene. *Cell* 43: 29-37.
- Kerr, W.G., G.P. Nolan, A.T. Serafini, and L.A. Herzenberg. 1989. Transcriptionally defective retroviruses containing *lacZ* for the in situ detection of endogenous genes and developmentally regulated chromatin. *Cold Spring Harbor Symp. Quant. Biol.* 54: 767-776.
- Kothary, R., S. Clapoff, A. Brown, R. Campbell, A. Peterson, and J. Rossant. 1988. A transgene containing *lacZ* inserted into the *dystonia* locus is expressed in neural tube. *Nature* 335: 435-437.
- Kozak, M. 1987. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* 15: 8125-8148.
- Maas, R.L., R. Zeller, R.P. Woychik, T.F. Vogt, and P. Leder. 1990. Disruption of formin encoding transcripts in two mutant *limb deformity* alleles. *Nature* 346: 853-855.
- Melton, D.A., P.A. Krieg, M.R. Rebagliata, T. Maniatis, K. Zinn, and M.R. Green. 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* 12: 7035-7056.
- Ohno, S. 1986. The total number of genes in the mammalian genome. *Trends Genet.* 2: 8.
- Okamoto, K., H. Okazawa, A. Okuda, M. Sakai, M. Muramatsu, and H. Hamada. 1990. A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell* 60: 461-472.
- O'Kane, C.J. and W.J. Gehring. 1987. Detection in situ of genomic regulatory elements in *Drosophila*. *Proc. Natl. Acad. Sci.* 84: 9123-9127.
- Reddy, S., J.V. DeGeorgi, H. von Melchner, and H.E. Ruley. 1991. Retrovirus promoter-trap vector to induce *lacZ* fusions in mammalian cells. *J. Virol.* 65: 1507-1515.
- Sambrook, J., E.F. Fritsch, and T. Maniatis, eds. 1989. *Molecular cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Schnieke, A., K. Harbers, and R. Jaenisch. 1983. Embryonic lethal mutation in mice induced by retrovirus insertion into the $\alpha 1(I)$ collagen gene. *Nature* 304: 315-320.
- Shedlovsky, A., T.R. King, and W.F. Dove. 1988. Saturation germ line mutagenesis of the murine *t* region including a lethal allele at the quaking locus. *Proc. Natl. Acad. Sci.* 85: 180-184.
- Skarnes, W.C. 1990. Entrapment vectors: A new tool for mammalian genetics. *Biotechnology* 8: 827-831.
- von Melchner, H., S. Reddy, and H.E. Ruley. 1990. Isolation of cellular promoters by using a retrovirus promoter trap. *Proc. Natl. Acad. Sci.* 87: 3733-3737.
- Weber, D.M., J. de Villiers, and W. Schaffner. 1984. An SV40 "enhancer trap" incorporates exogenous enhancers or generates enhancers from its own sequences. *Cell* 36: 983-992.
- Weither, H., T. Noda, D.A. Gray, A.H. Sharpe, and R. Jaenisch. 1990. Transgenic mouse model of kidney disease: Insertional inactivation of ubiquitously expressed gene leads to nephrotic syndrome. *Cell* 62: 425-434.

Skarnes et al.

- Wilkinson, D.G., S. Bhatt, P. Chavrier, R. Bravo, and P. Charnay. 1989. Segment-specific expression of a zinc-finger gene in the developing nervous system of the mouse. *Nature* 337: 461-464.
- Wilson, C., R.K. Pearson, H.J. Bellen, C.J. O'Kane, U. Grossniklaus, and W.J. Gehring. 1989. P-element-mediated enhancer detection: An efficient method for isolating and characterizing developmentally regulated genes in *Drosophila*. *Genes & Dev.* 3: 1301-1313.
- Woychik, R.P., R.L. Maas, R. Zeller, T.F. Vogt, and P. Leder. 1990. "Formins": Proteins deduced from the alternative transcripts of the *limb deformity* gene. *Nature* 346: 850-853.